基于多层特征融合可调监督函数卷积神经网络的人脸性别识别 *

石学超，周亚同†，池 越

(河北工业大学 电子信息工程学院 天津市电子材料与器件重点实验室，天津 300401)

摘 要：为了进一步提高性别识别的准确率，提出了一种基于多层特征融合与可调监督函数机制的结合的卷积神经网络（L-MFCNN）模型，并将之用于人脸性别识别。与传统卷积神经网络（CNN）不同，L-MFCNN 将多个浅层中间卷积层特征输出与最后卷积层特征输出相结合，融合多层卷积层的特征，不仅利用了深层卷积的整体语义信息，还考虑了浅层卷积的细节局部纹理信息，使得性别识别更加准确。此外 L-MFCNN 还引入具有可调目标监督函数机制的 Large-Margin Softmax Loss 作为输出层，利用其调节不同的间隔（margin）的机制来有效引导深层卷积网络学习，使得同种性别间的类内间距更小，不同性别间的类间距更大，获得更好的性别识别效果。在多个人脸数据集上的性别识别实验结果表明，L-MFCNN 的识别准确率要高于其他传统的卷积网络模型。L-MFCNN 模型也为将来的人脸性别识别研究提供了新的思路与方向。

关键词：人脸性别识别；多层特征融合；卷积神经网络；深度学习

中图分类号：TP391.41

Face gender recognition based on multi-layer feature fusion convolution neural network with adjustable supervisory function

Shi Xuechao, Zhou Yatong†, Chiyue

(*Tianjin Key Laboratory of Electronic Materials and Devices*, *School of Electronics & Information Engineering Hebei University of Technology*, *Tianjin* 300401, *China*)

Abstract: In order to further improve the accuracy of gender recognition, this paper proposed the convolution neural network model based on multi-layer feature fusion with adjustable supervisory function , L-MFCNN, then it used for face gender recognition. Unlike the traditional convolution neural network, L-MFCNN combined the output of multiple shallow convolution layers with the final convolution layer output. Fusion the characteristics of multi-layer convolutions, not only use the high-level semantic information, but also consider the bottom of the details of the texture information, making the face gender recognition more accurate. While using the Large-Margin Softmax Loss could adjust the margin function, it could explicitly encourages the same gender intra-class compactness and the different gender inter-class separability to get better face gender recognition. The face gender recognition experiment data on multiple face data sets show that the accuracy of L-MFCNN is higher than that of traditional convolution network. Besides, L-MFCNN also provides the new ideas and directions for the future gender recognition of face.

Key Words: face gender detection; multi-layer feature fusion; convolution neural network(CNN); deep learning

0 引言

人脸性别识别是人脸属性分析的关键步骤，是通过人脸图像信息自动发掘和分析人脸属性的二分类问题，已在视频监控、智能用户界面、人口统计等领域得到应用。人脸性别识别几乎涉及从模式识别到深度学习[1]的各种方法，如人工神经网络（ANN）、主成分分析（PCA）、贝叶斯决策、支持向量机、AdaBoost 算法以及卷积神经网络等，其中， Golomb 等人[2]通过训练一个两层的人工神经网络（ANN）,这是第一次将神经网络应用于性别识别，在一个小的人脸数据集上取得了很好的效果。Bruunelli 等人[3]通过采用一个三层的反向传播网络对不同分辨率的人脸图像进行性别识别，该网络在 30 张大小为8×8 的低分辨率测试图像上获得了 93%的准确率。Tamura 等人[4]提出使用极限学习进行人脸性别识别，取得了较好的识别准确率。以上使用全连接神经网络进行人脸性别识别的方法，忽略了图像中像素间的二维相关性，因此，分类准确率也不够高。

基金项目：国家自然科学基金资助项目（61401307）；河北省科学技术研究与发展项目（11213565）；河北省引进留学人员资助项目（CL201707）

作者简介：石学超（1991-），男，硕士研究生，主要研究方向为信息智能处理、深度学习；周亚同（1973-），男（通信作者），教授，博导，博士，主要研究方向为机器学习、模式识别、数字图像与视频处理（zyt@hebut.edu.cn）；池越（1977-），男，副教授，博士，主要研究智能信息处理、机器学习.

目前，深度学习在计算机视觉和物体识别取得了突破性的成果。越来越多的研究者将卷积神经网络算法引入到人脸识别相关领域中，如人脸检测[5~7]、关键点标定[8,9]、人脸识别[10~11]及人脸年龄估计[12,13]的研究。然而，目前基于卷积神经网络的人脸性别识别研究才刚刚开始，Verma 等人[14]曾使用一个 6 层的深度卷积网络进行人脸性别识别，获得了比以往传统方法更高的准确率；王济民等人[15]曾使用简单 2 层卷积网络进行人脸性别识别，该网络只是利用传统卷积网络的稀疏连接和权值共享的特性，性别识别准确率并不是很高；董兰芳等人[16]采用了深度卷积网络和随机森林相结合的方式进行人脸性别识别，该方法利用卷积网络提取特征，然后再使用随机森林分类器进行分类，虽然识别精度较高，但实现过于繁琐复杂。张婷等人[17]提出了一种 9 层的跨层连接卷积神经网络（CCNN）模型，该模型将中间一个池化层的输出跨过两个卷积层和全连接层相连接，但由于池化过程的特征损失，得到的分类结果并不明显优于传统的卷积网络。

本文提出了一种基于多层特征融合与可调监督函数机制卷积神经网络的人脸性别识别算法。本文算法可以认为是对传统卷积神经网络的改进，综合了利用浅层卷积与深层卷积的多层特征信息，将第 2 层卷积层、第 4 层卷积层与第 5 层卷积层的输出相结合，利用多层特征融合的方式来增强最后输入到全连接层图像语义特征信息，提高分类的准确率；此外，本文算法还引入了具有可调节机制的目标监督函数 Large-Margin Softmax Loss[18]，有效引导网络学习使得同类样本间的类内间距更小，不同类样本间的类间间距更大，同时利用此目标函数可以调节不同的间隔的机制，以防止网络训练的过拟合，进一步增强本文算法的鲁棒性。

本文的研究意义体现于：a）利用多层特征融合特征，构建多层卷积融合的卷积神经网络结构来增强人脸特征的提取过程，提高人脸性别识别的准确率；b）引入 Large-Margin Softmax Loss 代替传统的 Softmax loss 作为最后目标函数来有效地监督模型的训练过程，进而得到一个更加有区分性的性别识别模型。

1　多层特征融合卷积神经网络模型

1.1　卷积神经网络

卷积神经网络包括前向传播和反向传播，卷积层和采样层交替进行。卷积层后有一个下采样层来减少计算时间和建立空间与结构的不变性[19]。

前向传播是从输入参数到输出参数结果计算一次。上层的输出就是当前层的输入，然后通过激活函数，计算当前层的输出：

$$x^l = f\left(W^l x^{l-1} + b^l\right) \tag{1}$$

其中：l 代表层数；W 表示权值；b 是一个偏置；f 是激活函数。

反向传播就是从前向传播计算出的结果和给定样本的标签做误差计算。考虑平方差损失函数，对于包含 c 个类别，N 个训练样本的分类问题，误差函数表示为

$$E^N = \frac{1}{2}\sum_{x=1}^{N}\sum_{k=1}^{c}\left(l_k^x - y_k^x\right)^2 \tag{2}$$

其中：l_k^x 表示第 n 个样本对应的标签的第 k 维；y_k^x 表示第 n 个样本对应的网络输出的第 k 个输出。

反向传播会更新卷积层，上一层的特征映射和一个可以训练的核进行卷积运算，卷积运算的结果经激活函数后的输出形成了这一层的特征映射。每一层的输出映射可能与上一层的几个特征映射的卷积有关系。卷积层的一般形式为

$$x_j^l = f\left(\sum_{i=M_j} x_i^{l-1} * k_i^l + b_j^l\right) \tag{3}$$

其中：k 为卷积核；M_j 表示输入特征的一个选择；b 为一个偏置。

下采样操作并没有改变特征映射的数据，只是将特征映射的大小变小。如果采样算子大小为 $n \times n$，那么经过一次下采样，特征映射的大小变为原来特征的 $1/n$，表达式为

$$x_j^l = f\left(W_j^l down\left(x_j^{l-1}\right) + b_j^l\right) \tag{4}$$

其中：$down(.)$ 表示一个下采样函数。

1.2　权值更新

卷积神经网络常使用 BP 算法和梯度下降算法更新权值和偏置，梯度下降算法需要计算出损失函数对应节点的权值和偏导数，具体表达公式如下：

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial W_{ij}^{(l)}} J\left(W,b;x^{(i)},y^{(j)}\right)\right] + \gamma W_{ij}^{(l)} \tag{5}$$

$$\frac{\partial}{\partial b_{ij}^{(l)}} J(W,b) = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_{ij}^{(l)}} J\left(W,b;x^{(i)},y^{(i)}\right) \tag{6}$$

其中：$W_{ij}^{(l)}$ 为第 l 层的 (i,j) 节点权值；$b_{ij}^{(l)}$ 为第 l 层的 (i,j) 节点偏置；γ 为规则化参数。则更新后的反馈误差为

$$\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)}\right) \cdot f'\left(z_i^{(l)}\right) \tag{7}$$

$$\delta_i^{(n_l)} = -\left(y_i - a_i^{(n_l)}\right) \cdot f'\left(z_i^{(n_l)}\right) \tag{8}$$

其中：n_l 为最终输出层；$\delta_i^{(l)}$ 为第 l 层 i 节点残差；$z_i^{(l)}$ 为输入到第 l 层 i 节点的值；$a_i^{(n_l)}$ 为第 n_l 层 i 节点激活值；s_{l+1} 为第 $l+1$ 层节点数。

1.3　多卷积层特征融合理论分析

对于传统 CNN 就是把图像进行一层层映射与筛选的过程，映射到最后一层的特征就是最后提取的结果。在整个映射过程

中，对于不同的 CNN 层学习表现的特征是不同的，早期浅层的网络含有更多的是层次信息，如边缘纹理信息；网络的最后一层输出更抽象的语义信息。图 1 所示是对不同层的滤波器特征的可视化图像[30]。从插图中可以清晰地发现不同层的网络所提供的不同的信息。每一层的卷积在针对于相同的输入图像会体现不同的表征特征。此外，卷积层比全连接层表现出更多的空间信息。

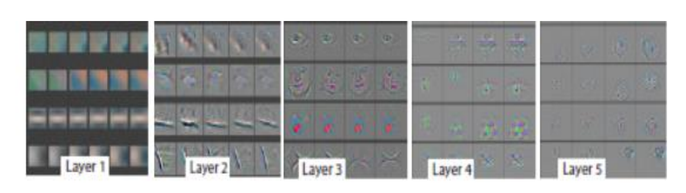


图 1　不同卷积层滤波器特征的可视化图像

另外，如图 1 所示，随着卷积层的不断加深，所体现的特征也在随之改变，由开始浅层明显的纹理信息变到了最后语义信息的集中；CNN 卷积的过程其实即不断刷选的过程，但在这个过程中，过滤掉的浅层的特征并不一定对最后的识别或分类是无用的，如浅层第一层、第二层的边缘纹理信息，也是含有一定信息的，对图像是具有一定的表能力的。因此本文考虑把浅层的卷积层的信息特征与深层的特征进行多层特征融合来增强整体模型对图像的表现能力。

1.4　多层特征融合卷积神经网络

对于传统的卷积神经网络来说，只是利用深层次的卷积特征信息来进行构建分类器，间接损失放弃了浅层次卷积层的细节纹理信息。针对上述问题，本文基于多层特征融合的思想，提出了一种基于多层特征融合卷积神经网络的人脸性别识别算法。此算法的深度卷积模型如图 2 所示。该模型包含 1 个输入层(data)、5 个卷积（Conv1,Conv2,Conv3,

Conv4,Conv5）、3 个池化层（Max1,Max2,Max3）、2 个上采样层（Upsampling1,Upsamoling2）、1 个融合层（Concat）、一个全连接层（FC）、一个输出层（Large-Margin Softmax Loss）；其中输入层输入图像信息，然后通过 5 个卷积层和 2 个池化层进行图像提取，分别对 Conv4、Conv5 的输出进行 2 倍上采样操作，得到与 Conv2 输出同样大小的特征图 Feature map，再将这三部分特征通过融合层（Concat）进行多层特征的融合处理，最后经全连接层(FC)对融合的特征进行分类送到输出层，输出层的两个节点分别代表输入图像所属的类别。

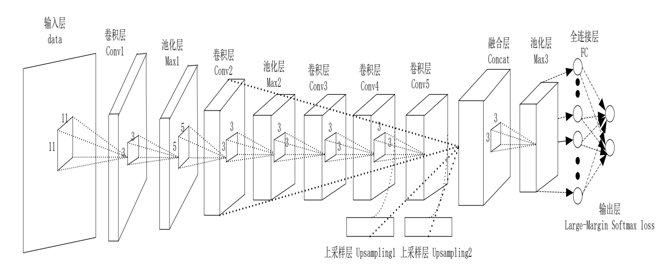


图 2　多层特征融合卷积神经网络结构示意图

对应多层特征融合卷积神经网络的网络各层的卷积核及滑动步长等参数设计如表 1 所示。

表 1　多层特征融合卷积神经网络的网络描述

Layer	Type	Kernel size	Stride	Pad
data	Input			
Conv1	Convolution	11×11	4	
Max1	Pooling	3×3	2	2
Conv2	Convolution	5×5	1	
Max2	Pooling	3×3	2	
Conv3	Convolution	3×3	1	1
Conv4	Convolution	3×3	1	1
Conv5	Convolution	3×3	1	1
Upsampling1	UpSampling			
Upsampling2	UpSampling			
Max3	Pooling	3×3	2	
Concat	Concat			
FC	Fully Connection			

1.5　输出层损失函数的选择

以往卷积网络的最常用输出层为交叉熵函数 Softmax Loss[20],通过定义第 i 个输出特征 X_i 以及它的标签 y_i 时，可得到输出层 softmax loss 的表达式为

$$L=\frac{1}{N}\sum_i L_i=\frac{1}{N}\sum_i -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \tag{9}$$

其中：f_j 表示最终全连接层的类别输出向量 f 的第 j 个元素；N 为训练样本的个数。由于 f 是全连接层的激活函数 W 的输出，所以 f_{y_i} 可以表示为 $f_{y_i}=W_{y_i}^T X_i$,最终的损失函数表示为类间角度 θ（$0\le\theta_j\le\pi$）的表达式：

$$L_i=-\log\left(\frac{e^{\|W_{y_i}\|\|X_i\|\cos(\theta_{y_i})}}{\sum_j e^{\|W_j\|\|X_i\|\cos(\theta_j)}}\right) \tag{10}$$

虽然 softmax 在深度卷积网络中有着很广泛的应用，但是这种形式并不能够有效地学习使得类内较为紧凑、类间较为离散的特征，softmax 的目的是使得 $W_1^T x > W_2^T x$，即通过不等式 $\|W_1\|\|x\|\cos(\theta_1) > \|W_2\|\|x\|\cos(\theta_2)$ 来得到 x 的正确分类结果，而 Large-Margin Softmax Loss 的监督函数[18]优势就是通过增加一个正整数变量 m，从而产生一个决策余量，进而更加严格地约束上述不等式，即

$$\|W_1\|\|x\|\cos(\theta_1)\ge\|W\|\|x\|\cos(m\theta_1)>\|W_2\|\|x\|\cos(\theta_2) \tag{11}$$

其中：$0\le\theta_1\le\frac{\pi}{m}$。如果 W_1 和 W_2 能够满足 $\|W_1\|\|x\|\cos(m\theta_1)>\|W_2\|\|x\|\cos(\theta_2)$，那么就必满足 $\|W_1\|\|x\|\cos(\theta_1)>\|W_2\|\|x\|\cos(\theta_2)$。这样的约束对学习 W_1 和 W_2 的过程提出了更高的要求，从而使得 1 类和 2 类有了更宽的分类决策边界。Large-Margin Softmax Loss 的表达式为

$$L_i = -\log\left(\frac{e\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i})}{\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i}) + \sum_{j\neq y_i} e\|W_j\|\|x_i\|\cos(\theta_j)}\right) \quad (12)$$

其中：$\psi(\theta)$可以表示为

$$\psi(\theta)=\begin{cases}\cos(m\theta), 0\le\theta\le\dfrac{\pi}{m}\\ \cos\left(\dfrac{\pi}{m}\right), \dfrac{\pi}{m}<\theta\le\pi\end{cases} \quad (13)$$

从而可以通过调节不同的间隔 m 的值来调节分类的边界，通过调节学习的难度来预防网络训练的过拟合现象。

2 实验结果与分析

为验证本文 L-MFCNN 模型的识别性能，本文将其与传统的 CNN 以及张婷等人提出的跨连卷积神经网络（CCNN）模型的性别识别性能进行了对比。实验共采用 6 个人脸数据集进行分析比较，分别为 AR（Aleix Martinez and Robort Benavente）数据集[24]、ORL（Olivetti Research Laboratory）数据集[25]、UMIST（University of Manchester Institute of Science and Technology）数据集[26]、FERET(Face recognition technology)数据集[27]、LFW(Labeled faces in the wild)数据集[28]、CelebFace 数据集[29]。在 GPU 1.2 GHz，显存 12 GB 的 WinFast gs4800 服务器上进行训练，在 CPU 3.6 GHz, 内存 8 GB 的预装 Windows 10 旗舰版 64 位操作系统的个人计算机上进行测试，深度学习框架为 Caffe，软件编程环境为 Python2.7.5，MATLAB2014a。

本文算法 L-MFCNN 网络与传统的 CNN 网络均采用五层卷积的 AlexNet 网络结构，并且每层的参数保持一致；激活函数采用 ReLU 激活函数[25]：

$$f(z)=\max(0,z) \quad (14)$$

ReLU 可由 $f(x)=\log(1+e^z)$逼近，另外惩罚系数取值 $\lambda=0.000001$，动量项系数设置为 $\beta=0.9$ 初始学习率为 $lr=0.01$。UMIST 和 FERET 数据集上的最大迭代次数设为 10 000 次，每 100 次保存一个模型 model；在 LFW、CelebFace 数据集上的最大迭代次数设为 100 000 次，每 1 000 次保存一个模型 model。

2.1 实验室数据制作

AR 集、ORL、UMIST 和 FERET 数据集为较少的数据集，四种数据集均转换为灰度图像，LFW、CelebFace 数据集为大数据集，对于 LFW、CelebFace 数据大的数据集采用 RGB 彩色图像进行训练，所有数据集图像转换到128×128。六种数据集的训练与测试样本的分配如表 2 所示,另外每个数据集的部分示例图像如图 3 所示。

2.2 实验测试分析

为了验证采用多层特征融合方式和引入的 Large-Margin Softmax Loss 作为输出层对性别识别性能的影响，实验分为四组进行，第一组是对 L-MFCNN 模型各卷积层的特征图可视化实验；第二组是只采用多层特征融合方式进行数据的训练与测试；第三组是在采用多层特征融合方式的基础上，另外将传统的 softmax loss 换为 Large-Margin Softmax Loss 来进行对比验证；第四组是对 L-MFCNN 模型在各数据集的男性和女性识别情况进行测试实验分析。

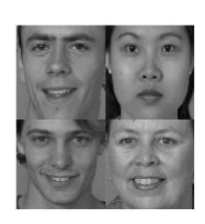
表 2 各数据集的训练和测试样本分配

数据集	训练集			测试集		
	男	女	混合	男	女	混合
AR	1100	1100	2200	200	200	400
ORL	320	30	350	40	10	50
UMIST	230	50	280	40	20	60
FERET	620	500	1120	150	123	273
LFW	8000	2500	10500	2000	500	2500
CelebFace	12000	5000	17000	2000	1000	3000


(a) AR 数据集


(b) ORL 数据集


(c) UMIST 数据集


(d) FERET 数据集




(e) LFW 数据

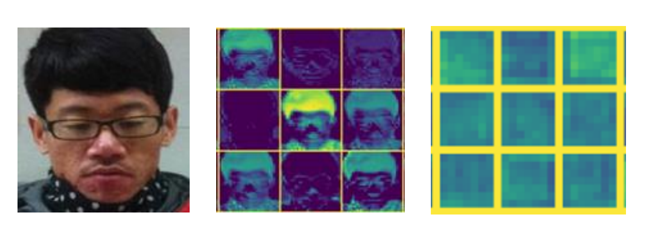
(f) CelebFace 数据集

图 3 六个数据集部分示例图像

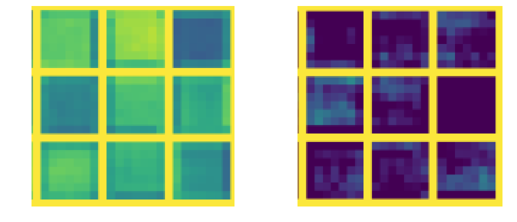
2.2.1 L-MFCNN 模型卷积层可视化分析

第一组实验：通过对多层特征融合模型 L-MFCNN 训练好的模型进行卷积层输出特征图 feature map 可视化图像进行了分析验证，实验中分别把卷积层 Conv2、Conv4、Conv5 对应输出的特征图进行了可视化呈现，另外将以上三层特征融合后的最后输出的特征图进行了可视化操作，对应卷积层特征图可视化结果如图 4 所示。

从图中可以清晰分析得到：对于浅层 Conv2 卷积层输出的特征图含有原始图像较强的细部纹理边缘信息；再到较深的 Conv4 层输出的特征图，体现的图像的语义信息在增多；到深层的 Conv5 卷积层输出的特征图很难看出原始图的边缘纹理信息了，体现的是深层卷积的整体语义信息；但当把卷积层 Conv2,Conv4,Conv5 三层输出的特征融合后，输出的图像如图 4（f）所示，此特征图即含有浅层图像的细部纹理信息，又含有深层卷积层图像整体语义信息，加强了模型对图像的辨识能力，则可以有效地对人脸的性别图像进行区分。



(a)原始图像　(b)Conv2 层可视化图像　(c)Conv4 层可视化图像



(e) Conv5 层可视化图像　(f) Conv2Conv4Conv5 三层融合可视化图像

图 4　L-MFCNN 模型卷积层可视化结果

2.2.2 多层特征融合实验对比分析

第二组实验：通过对只进行多层特征融合模型(MFCNN)与传统的 CNN、跨连卷积神经网络（CCNN）模型进行对比实验，三种模型均是以采用五层卷积的 AlexNet 网络结构为基础。传统的 CNN 保持原始网络结构不变；跨层连接神经网络（CCNN）以按第二个隐藏层（第一个池化层）与最后全连接阶层连接的方式进行跨层连接；而本文的多层特征融合模型以 2.2 节提到的将三个卷积层（Conv2、Conv4、Conv5）提取的特征接着传给融合层（Concat）进行多层特征的融合。三种模型在 6 个人脸数据集的性别识别结果如图 5 所示。

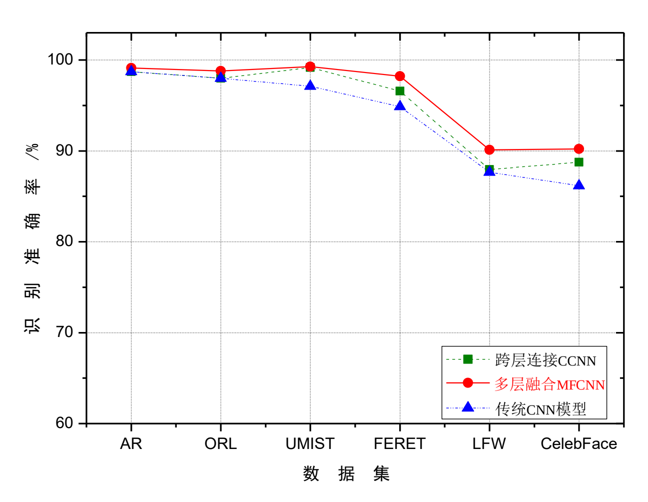


图 5　单独多层特征融合模型识别准确率对比

从图 5 可以看出，多层特征融合 MFCNN 模型在 AR、ORL、UMIST、FERET、LFW、CelebFace 6 个数据集上的识别准确率均要高于 CNN 与 CCNN；MFCNN 模型在小的数据集 AR、UMIST 上分别达到了 99.13%、99.28%的识别率，在相对较大的数据集 FERET、LFW、CelebFace 上识别准确率分别为 98.21%、90.16%、90.08%，相对于跨层连接 CCNN 模型分别提高了 1.63%、2.2%、1.33%；跨层连接模型 CCNN 只是利用两个池化层提取的特征去识别分类，并未增加太多原始图像的特征信息，识别准确率提升并不是很大；而 MCCNN 是对三个卷积层 Conv2、Conv4、Conv5 提取的特征进行了融合，即利用了深层卷积层 Conv4、Conv5 的输出图像的整体语义信息特征，还同时考虑了浅层卷积层 Conv2 的细节局部纹理信息特征，从而获得了较好的分类性能。

2.2.3 多层特征融合加入 Large-Margin Softmax Loss 监督函数实验对比分析

第三组实验：在第一组实验的基础上通过在多层特征融合模型引入了 Large-Margin Softmax Loss 输出层构成 L-MFCNN 模型与传统的 CNN、跨连卷积神经网络 CCNN 模型进行对比实验，三种模型均是以采用五层卷积的 AlexNet 网络结构为基础。传统的 CNN 保持原始网络结构不变；跨层连接神经网络（CCNN）以按第二个隐藏层（第一个池化层）与最后全连接阶层连接的方式进行跨层连接；而 L-MFCNN 模型以 2.2 节提到的将三个卷积层 Conv2、Conv4、Conv5 提取的特征接着传给融合层（Concat）进行多层特征的融合，最后用 Large-Margin Softmax Loss 作为输出层进行性别间类别监督，通过调节不同的间隔 m 的值，来调节分类的边界进而来获取较高的性别识别准。五 种模型在 6 个人脸数据集的性别识别结果如图 6 所示。

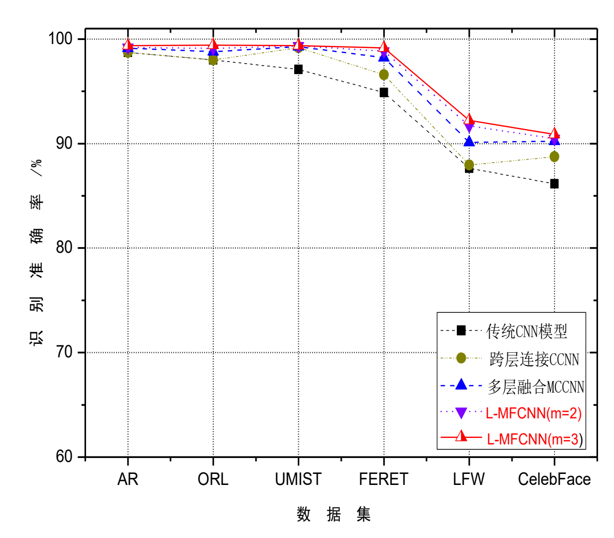
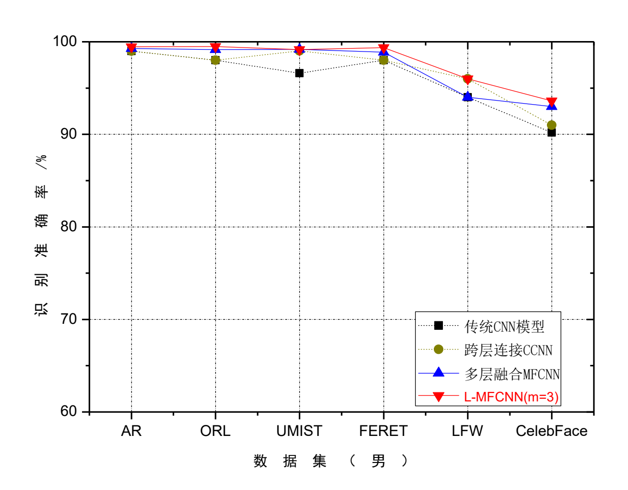


图 6　多层特征融合加入 Large-Margin Softmax Loss 实验对比

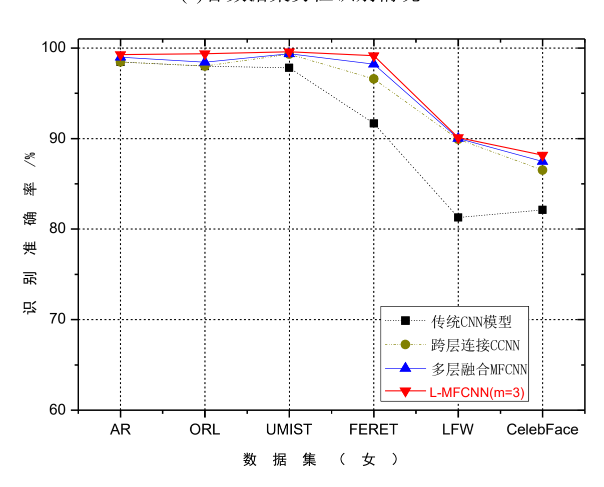
对于 L-MFCNN 模型依靠 Large-Margin Softmax Loss 输出层作为监督函数，依次调节不同的间隔 m 的值来产生不同决策余量，进而获得更宽的分类决策边界，实验中当 m 取 2、3 时效果最佳。从图 4 可以看出，当 m 取 3 时 L-MFCNN 模型在 AR、ORL、UMIST、FERET、LFW、CelebFace 6 个数据集上的识别准确率均要高于其他各种模型识别准确率，L-MFCNN 模型在小的数据集 AR、ORL、UMIST 上分别达到了 99.36%、99.42%、99.38%的识别率，相对于单独特征融合 MFCNN 识别率依次提高了 0.17%、0.31%、0.03%、在相对较大的数据集 FERET、LFW、CelebFace 上识别准确率分别为 99.12%、92.22%、90.88%，相对于 MFCNN 模型要分提高了 0.27%、0.54%、0.42%；可见 L-MFCNN 相比于多层融合 MFCNN 获得了更好的性别识别性能。

2.2.4 L-MFCNN 在各数据集男性和女性识别情况分析

第四组实验：为了比较和说明 L-MFCNN、单独多层融合 MFCNN、跨层连接模型 CCNN 在 6 个数据集中男性和女性的识别效果，把每个数据集的测试集男性和女性图像分别挑选了出来，分别用第二组实验中以上四种模型训练好的卷积模型进行了测试，测试结果如图 7 所示。



(a)各数据集男性识别情况



(b)各数据集女性识别情况

图 7　测试结果

从图 7 可以看出，本文的 L-MFCNN 模型在调节间隔 m 为 3 时，在 AR、ORL、UMIST、FERET、LFW、CelebFace 6 个数据集男性和女性测试分类准确率都不低于 CNN、CCNN 和 MFCNN 模型。在相对较大的数据集 FERET、LFW、CelebFace 上男性分类准确率分别为 98.88%、90.16%、90.08%，与传统 CNN 相比分别高出 1.36%、2.02%、3.3%；另外在女性测试集上每个模型的识别性能要相对于男性识别要低一些，因为在三个数据集 FERET、LFW、CelebFace 上女性训练的样本要明显少于男性的训练的样本，进而使得在卷积网络训练过程中男性这一类学习的性对于好些。

3　结束语

为了进一步提高性别识别的准确率，本文提出了一种多层特征融合与可调目标监督函数机制的 Large-Margin Softmax Loss 结合的卷积神经网络模型 L-MFCNN。该模型将多个浅层中间卷积层特征输出与最后卷积层特征输出相结合，融合多层卷积层的特征，通过利用深层卷积的整体语义信息特征和浅层卷积层的细节局部纹理信息特征，获得更加准确的性别识别结果。此外，还引入了具有可调目标监督函数机制的 Large-Margin Softmax Loss 作为输出层，利用其调节不同的间隔的机制来有效地引导深层卷积网络学习使得同种性别间的类内间距更小，不同性别间的类间间距更大，获得更好识别分类效果。实验结果表明，在 6 个人脸数据集上 L-MFCNN 模型识别准确率均高于其他模型。

参考文献：

[1] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7533): 436-444.

[2] Golomb B A, Lawrence D T, Sejnowski T J. SexNet: a neural network identifies sex from human faces [C]// Advances in Neural Information Processing Systems. 1991: 572-579.

[3] Brunelli R, Poggio T. HyberBF networks for gender classification [C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. 2007.

[4] Tamura S, Kawai H, Mitsumoto H. Male//female identification from 8× 6 very low resolution face images by neural network [J]. Pattern recognition, 1996, 29 (2): 331-335.

[5] Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2002: 130-136.

[6] Farfade S S, Saberian M J, Li L J. Multi-view face detection using deep convolutional neural networks [C]// Proc of the 5th ACM on International Conference on Multimedia Retrieval. New York: ACM Press, 2015: 643-650.

[7] Liao S, Jain A K, Li S Z. A fast and accurate unconstrained face detector [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2016, 38 (2): 211-223.

[8] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499-1503.

[9] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57 (2): 137-154.

[10] Sun Yi, Liang Ding, Wang Xiaogang, et al. DeepID3: face recognition with very deep neural networks [J]. Computer Science, 2015.

[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015: 3431-3440.

[12] Levi G, Hassner T. Age and gender classification using convolutional neural networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015: 34-42.

[13] Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent age

from a single image without facial landmarks [J]. International Journal of Computer Vision, 2016: 1-14.

[14] Verma A, Vig L. Using convolutional neural networks to discover cogntively validated features for gender classification [C]// Proc of IEEE International Conference on Soft Computing and Machine Intelligence. 2014: 33-37.

[15] 汪济民，陆建峰. 基于卷积神经网络的人脸性别识别 [J]. 现代电子技术, 2015, 38 (7): 81-84.

[16] 董兰芳，张军挺. 基于深度学习和随机森林的人脸年龄和性别分类研究 [J//OL]. 计算机工程, : 1-6 (2017-05-23) .

[17] 张婷，李玉鑑，胡海鹤，等. 基于跨连卷积神经网络的性别分类模型 [J]. 自动化学报, 2016, 42 (6): 858-865.

[18] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks [C]// Proc of International Conference on International Conference on Machine Learning. 2016: 507-516.

[19] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[20] Shen W, Wang X, Wang Y, et al. Deepcontour: a deep convolutional feature learned by positive-sharing loss for contour detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3982-3991.

[21] Mansanet J, Albiol A, Paredes R. Local deep neural networks for gender recognition [J]. Pattern Recognition Letters, 2016, 70: 80-86.

[22] Hassner T, Harel S, Paz E, et al. Effective face frontalization in unconstrained images [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4295-4304.

[23] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks [C]// Proc of the 14th International Conference on Artificial Intelligence and Statistics. 2011: 315-323.

[24] Sanguansat P. Face hallucination using bilateral-projection-based two-dimensional principal component analysis [C]// Proc of IEEE International Conference on Computer and Electrical Engineering. 2008: 876-880.

[25] Hightower J, Borriello G. Location systems for ubiquitous computing [J]. Computer, 2001, 34 (8): 57-66.

[26] Shen X, Lin Z, Brandt J, et al. Detecting and aligning faces by image retrieval [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3460-3467.

[27] Phillips P J, Wechsler H, Huang J, et al. The FERET database and evaluation procedure for face-recognition algorithms [J]. Image and vision computing, 1998, 16 (5): 295-306.

[28] Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report 07-49 [R]. Amherst: University of Massachusetts, 2007.

[29] Guo Y, Zhang L, Hu Y, et al. Ms-celeb-1m: a dataset and benchmark for large-scale face recognition [C]// Proc of European Conference on Computer Vision. Springer International Publishing. 2016: 87-102.

[30] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [J]. Computer, 2013, 8689: 818-833.